

**Request to Archive
With The National Centers for Environmental Information
For International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Release 3.0 Raw
Individual Reports in the International Maritime Meteorological Archive Format version 1
(IMMA1)
Provided by NCAR**

2016-05-03

This information will be used by NCEI to conduct an appraisal and make a decision on the request.

1. Who is the primary point of contact for this request?

Eric Freeman
NCDC/GCAD/IAB
Marine Observation Analyst
828-271-4463
eric.freeman@noaa.gov

2. Name the organization or group responsible for creating the dataset.

DOC/NOAA/NESDIS/NCEI > National Centers for Environmental Information, NESDIS, NOAA, U.S. Department of Commerce

3. Provide an overview summarizing the scope of data you want to archive. Describe the outputs, data variables, including their measurement resolution and coverage.

UPDATE:

2 new data types need to be added to this archive. In addition to the IMMA files, monthly summary group statistics are also produced. These need to be archived at NCEI to aid in the transition of duties from NOAA/ESRL to NCEI and set a baseline from where future NRT and ICOADS processing will start.

The MSG files are in a packed binary format and netCDF format. These files will be archived for internal use.

The International Comprehensive Ocean-Atmosphere Data Set (ICOADS) is the largest surface marine archive in the world. It contains marine reports from 1662 to present and underpins many NCEI and international climate products.

ICOADS Release 3.0 (R3.0) is the latest full update to the marine archive, and the first since 2007. With 455 Million records, it continues to be the best available source for global marine observations.

ICOADS full releases are currently archived in DSI 1173_02 for internal use. This set will need to be added to that collection for internal access from the HDSS by ICOADS, ERSST and OISST team members. Additional users within NCEI (i.e. World Ocean Database, and others) may need to access the data. However, these raw files are not for public consumption and will not be accessed from the HDSS but on rare occasions.

The dataset has many upgrades from previous versions including: a) new data sources, b) improved data format, and c) general improvements e.g. removal/correction of erroneous data/data sources.

4. What is the time period covered by the dataset? (YYYY-MM-DD, YYYY-MM or YYYY)

From 1662-10-01 to 2014-12-31

5. Edition or version number(s) of the dataset:

Release 3.0 (R3.0)

6. Approximate date when the dataset was or will be released to the public:

2016-04-29

7. Who are the expected users of the archived data? How will the archived data be used?

This will be a deep archive of the full, raw release and will only be needed internally on rare occasions. This set will also be processed through some additional ICOADS qc to be served to the public and archived in associated DSI 1173_01.

8. Has the dataset undergone user evaluation and/or an independent review process? Did NCEI participate in design reviews?

Data is currently being validated by International Partner organizations using the data in mock runs of climate products and is also being tested internally by ERSST.

9. Describe the dataset's relationship to other archived datasets, such as earlier versions or related source data. If this is a new version, how does it improve upon the previous version(s)?

These data are highly associated with data in the following datasets:

1173_01 SA and metadata records C00685 (NCDC Global Marine Obs in IMMA) & C00686(ICOADS-ET).

1173_02, no SA, C00606 (ICOADS 2.5)

1178_01, SA and C00872 (International Comprehensive Ocean-Atmosphere Data Set (ICOADS) Near-Real-Time (NRT))

10. List the input datasets and ancillary information used to produce the data.

Input data sets are archived in DSI 1179_01, approximately 1 Billion input records.

11. List web pages and other links that provide information on the data.

<http://icoads.noaa.gov/>

<http://icoads.noaa.gov/r3.html>

<http://icoads.noaa.gov/e-doc/msg>

http://www.esrl.noaa.gov/psd/data/gridded/conventions/cdc_netcdf_standard.shtml

12. List the kinds of documents, metadata and code that are available for archiving. For example, data format specifications, user guides, algorithm documentation, metadata compliant with a standard such as ISO 19115, source code, platform/instrument metadata, data/process flow diagrams, etc.

1. Format Documentation (being finalized)

Processing codes (duplicate elimination, qc)

13. Indicate the data file format(s).

1. IMMA

2. netCDF-4

14. Are the data files compressed?

gzip

15. Provide details on how the files are named and how they are organized (e.g., file_name_pattern_YYYYMM.tar in monthly aggregations).

TBD - we are on Beta version #3 and will get final file names in the next couple of weeks

16. Explain how to access sample data files and/or a file listing for previewing. If it is not available now, when will it be available?

I can provide those when they are available. Currently working with partial records for ease of transfer and testing.

17. What is the total data volume to be submitted?

Historic Data: all historic data or data submitted as a completed collection.

Total Data Volume: 367GB

Number of Data Files: 114200

18. Are later updates, revisions or replacement files anticipated? If so, explain the conditions for submitting these additional data to the archive.

Eventually a new release will be produced, but that is not in the near-term. However, updates to the R3.0 data provided to the archive are not anticipated to change.

19. Describe the server that will connect to the ingest server at NCEI for submitting the data.

Physical Location: NOAA/ESRL, Boulder, CO

System Name: Sandra Lubker iMac Computer

System Owner: NOAA/ESRL

Additional Information: IP Address needs to be determined.

20. What are the possible methods for submitting the data to NCEI? Select all that apply.

1. FTP PUSH
2. SFTP PUSH

21. Identify how you would like NCEI to distribute the data. Web access support depends on the resources available for the dataset.

1. No web access

22. Will there be any distribution, usage, or other restrictions that apply to the data in the archive?

Constraint Type	Description
Access	Internal-only access

23. Discuss the rationale for archiving the dataset and the anticipated benefits. Mention any risks associated with not archiving the dataset at NCEI.

This is one of the most widely used and cited datasets in existence. Benefits are wide as stakeholders span many scientific and government decision-making groups and disciplines.

24. Are the data archived at another facility or are there plans to do so? Please explain.

Data will also be archived at NCAR, but it's not a national archive. As a longstanding partner, they will mirror the data we hold at NCEI.

25. Is there an existing agreement or requirement driving this request to archive? Have you already contacted someone at NCEI?

Being the lead for ICOADS, NCEI should archive this data as it is widely used internally and externally.

26. Do you have a data management plan for your data?

This release will complete our near-term deliverables and drafting of a 10 year workplan will be started soon.

27. Have funds been allocated to archive the data at NCEI?

Base funded project

28. Identify the affiliated research project, its sponsor, and any project/grant ID as applicable.

N/A

29. Is there a desired deadline for NCEI to archive and provide access to the data?

Archive by: 2016-05-26

Accessible by: 2016-06-20

30. Add any other pertinent information for this request.

Updated information to include 2 new data types. Number of files and full storage size updated to reflect all data files in this collection.